四个竹秆变异毛竹变型的全基因组序列分析

牟少华[1*]，李娟[1]，李雪平[1]，高健[1]
（1. 国际竹藤中心 竹藤科学与技术重点实验室，北京 100102）

摘要：毛竹是我国重要的经济竹种，在长期栽培适应过程中产生了丰富的变异。为揭示毛竹竹秆变异变型的全基因组突变类型，以黄皮毛竹、金丝毛竹、绿皮花毛竹和花毛竹 4 个毛竹变型为实验材料，采用高通量重测序技术获得全基因组序列，进行单核苷多态性（SNP）、小片段插入缺失（InDel）和结构变异（SV）检测和注释，并将变异基因进行功能注释。结果表明：花毛竹基因组检测得到的基因变异数最多，为 12 555 个；金丝毛竹样品变异位点数最少，为 11 923 个；4 个样品都有 7 000 多个变异基因得到功能注释。GO 注释分类包括细胞组件、分子功能和生物过程三个基因功能分类体系的 56 个功能组。在细胞组分方面，叶绿素合成相关基因有 2 431 个；在生物过程方面，参与类胡萝卜素合成过程的基因有 75 个，参与花青素合成过程中的调控以及紫外光下组织中花青素积累的相关基因有 80 个。COG 分类表明参与复制、重组和修复的基因数为 369 个，信号转导机制的基因数为 291 个，转录的相关基因 222 个。通过 KEGG 数据库系统地分析变异基因参与的黄酮类、类胡萝卜素等物质代谢合成途径。深入研究这些差异基因的调控途径，从 DNA 水平上解释竹秆的变异机制，可以为深入研究毛竹种内丰富的多态性和遗传变异提供数据支持，阐析不同变异类型的基因家族、功能基因等遗传基础。

关键词：毛竹，变型，全基因组重测序，基因注释
中图分类号：S795.7　　**文献标识码：**A　　文章编号：

Genomic sequence analysis of four culm variants of Moso bamboo (*Phyllostachys edulis*) on culm

MU Shaohua[1*], LI Juan[1], LI Xueping[1], GAO Jian[1]
（1. *Key Laboratory of Bamboo and Rattan Science and Technology, International Center for Bamboo and Rattan*, Beijing 100102, China）

Abstract: As an important and economic bamboo species in China, Moso bamboo has performed lots of variations with long-term cultivation conditions. For an overall understanding of the whole genome of four representative culm variations, *Phyllostachys edulis* f. *holochrysa*, *P. edulis* f. *gracilis*, *P. edulis* f. *nabeshimana* and *P. edulis* f. *huamozhu*, re-sequencing was used for high-throughput sequencing to detect its variations by molecular data. The single nucleotide

收稿日期：2021-01-14
基金项目：国家林草局“948”项目（2015-4-13）；国际竹藤中心基本科研业务费项目（1632016003）[Supported by Project‘948’ of the National Forestry and Grass Administration (2015-4-13); Basic Scientific Research Program of ICBR (1632016003)]。
作者简介：牟少华(1976 -)，博士，副研究员，主要从事竹类植物种质资源与遗传多样性研究，（E-mail）mush@icbr.ac.cn。
[*]通信作者

polymorphism (SNP), insertion-deletion (InDel) and structure variation (SV) were detected and annotated, and the mutant genes were compared with the functional databases. The results were as follows: *Phyllostachys edulis* f. *gracilis* had the lowest number of mutation sites, that was 11 923, and *Phyllostachys edulis* f. *huamozhu* had the highest number 12 555, of which more than 7 000 mutant genes were annotated. GO annotation classification included 56 functional groups of three functional classification systems: cellular component, molecular function and biological processes. In terms of cell components, there were 2 431 genes related to chlorophyll synthesis. In terms of biological processes, there were 75 genes involved in the synthesis of carotenoids and 80 ones involved in the regulation of anthocyanin synthesis and anthocyanin accumulation in tissues under ultraviolet light. COG classification showed that 369 genes involved in replication, recombination and repair, 291 ones in signal transduction mechanism, and 222 ones in transcription. The metabolic pathways of flavonoids, carotenoids and other substances involved in the mutant genes were analyzed by KEGG database. In-depth study of the regulatory pathways and interpretation of the variation mechanism on culm from the DNA level, can provide a data basis for further exploration of the rich polymorphism and genetic variation of Moso bamboo, and elucidate the genetic basis of gene family and functional genes of different variation types.

Key words: Moso bamboo (*Phyllostachys edulis*), variant, whole genome re-sequencing, gene annotation

毛竹 (*Phyllostachys edulis*) 是我国独有的传统经济竹种（江泽慧，2002），分布范围广阔，现有毛竹变型20多种（马乃训等，2014）。毛竹变型在形态上表现出丰富的多态性，尤其是秆色性状方面差异表现显著，例如花毛竹秆为黄色，有宽窄不等的绿色纵条纹，而绿皮花毛竹秆为绿色，但节间有淡黄色细纵条纹。秆色的变异大大丰富了园林观赏种类，提高了园林观赏价值。毛竹竹秆性状的遗传变异是遗传育种工作关注的重点。

目前，高通量测序技术可以分析一个物种的基因组的全貌，已经在谷子 (Bai et al., 2013; 贾小平等，2019)、水稻 (Takagi et al., 2013)、大豆 (Qi et al., 2014; Zhou et al., 2015; 张彦威等，2016)、黄秋葵（张少平等，2017）、厚朴（尹彦棚等，2020）、菜豆 (Jeremy et al., 2010) 和番茄 (Lin et al., 2014)等植物上得到了广泛应用。

近年来，随着分子生物学和组学技术的发展，毛竹全基因组序列获得公布 (Peng et al., 2013)，一些与毛竹性状相关的基因家族，如*AP2/ERF* (Wu et al., 2015)、*SAUR* (Bai et al., 2016)、*AQP* (Sun et al., 2016)、*SBP*-like (Pan et al., 2016)、*HD-Zip* (Chen et al., 2017)、*Hsp* (Xie et al., 2019)、*CO-Like* (Liu et al., 2016)等已进行鉴定和功能验证，但是在基因组层面上的研究较少，特别是毛竹变型秆色相关的研究薄弱，仅对黄槽毛竹和黄皮花毛竹两个毛竹变型基因组序列变异进行初步探索 (牟少华等，2020)，这在一定程度上限制了毛竹遗传育种的应用发展。从基因水平上揭示毛竹的变异程度，是分析毛竹变型形态差异产生原因的重要手段之一。因此，开展毛竹变型基因组研究，揭示毛竹变型全基因组突变类型，探究黄酮类和硝酸还原酶等代谢途径相关基因，对解析毛竹丰富的遗传多样性以及性状相关的遗传变异具有重要意义。

本研究以黄皮毛竹等 4 个有代表性的竹秆变异毛竹变型为研究对象，毛竹全基因组作为参考基因组，采用高通量测序技术，构建全基因组数据库，并利用生物信息学的方法对获得的核酸序列组装，检测并注释其单核苷酸多态性(SNP)、结构变异(SV)和小片段插入缺失(Indel)等，注释变异基因功能，积累基因组序列数据，以便为从全基因组水平上深入地分析毛竹的遗传变异，为遗传育种提供遗传基础。

1 材料与方法

1.1 材料

供试材料选自国际竹藤中心安徽太平试验中心种质资源圃。选取4个毛竹变型（表1）适量新鲜幼嫩的叶片，经液氮罐中速冻后，放入-80℃冰箱冻存备用。

表 1 四个毛竹变型样品简表

Table 1 Brief introduction of four variations of Moso bamboo

编号 ID	变型 variation	拉丁名 Latin name	形态特征 Morphological characteristics
R01	黄皮毛竹	*Phyllostachys edulis* f. *holochrysa*	秆和枝金黄色 Culms and branches golden yellow
R02	金丝毛竹	*P. edulis* f. *gracilis*	竹株小，秆壁较厚，基部的节间较长 Shorts height, culm walls thick, basal internodes longer
R03	绿皮花毛竹	*P. edulis* f. *nabeshimana*	秆绿色，但节间有淡黄色细纵条纹 Culms green, internodes with light yellow strips
R04	花毛竹	*Phyllostachys edulis* f. *huamozhu*	黄秆，有宽窄不等的绿色纵条纹 Culms yellow with green strips

1.2 方法

1.2.1 基因组测序

毛竹变型叶片 DNA 提取(Zidani et al., 2005)后，经过打断、损伤修复及连接接头、PCR 富集、文库质量检测，建成测序文库，在 Illunima Hiseq 2500 测序平台上运行获得原始数据，将数据过滤后得到高质量数据。

1.2.2 比对统计

使用 bwa 软件 (Li& Durbin, 2009) 将测数据序比对定位到已测序的毛竹基因组的位置，统计测序深度和基因组覆盖度等信息。

1.2.3 检测 SNP、InDel 和 SV

使用 Picard 软件 (Gordon et al., 2012) 去重复和 GATK 软件 (Mckenna et al., 2010) 预处理后，检测 SNP 和 InDel 变异。使用 BreakDancer 软件 (Chen et al., 2009) 检测 SV 变异，具体方法参照文献（牟少华等，2020）。

1.2.4 注释 SNP、InDel 和 SV

运用 SnpEff 软件 (Cingolani et al., 2012)注释 SNP、InDel 和 SV，具体方法参照文献（牟少华等，2020）。

1.2.5 注释功能基因

运用 BLAST 软件，对筛选得到的功能可能变异基因的基因序列与 GO (Ashburner et al., 2000)、COG (Tatusov et al., 2000)和 KEGG (Minoru et al., 2004) 等三大功能数据库，进行 BLAST 比对，得到基因注释。

2 结果与分析

2.1 与毛竹基因组比对

4个竹种通过高通量测序得到测序数据。金丝毛竹样品(R02)过滤后的Clean reads最少，为82 276 884 bp；花毛竹样品 (R04) 的Clean reads最多，为112 054 728 bp。定位到毛竹参考

基因组的占所有Clean reads数的百分比在99.45 %以上，双端均定位到毛竹参考基因组上并且距离符合测序片段的长度分布的占所有Clean Reads数的百分比在88 %左右，说明参考基因组选择合适，且相关实验过程不存在污染，测序reads的比对率会高于70%。另外，比对率的4个毛竹变型与毛竹参考基因组亲缘关系较近、基因组组装质量高，而且reads测序质量高。4个样品平均覆盖深度均在10×左右（表2）。

表2 四个样品数据产出统计表

Table 2 Output statistics among four samples

编号 ID	过滤后的 Clean_reads	定位比 Mapped (%)	双端定位比 Properly _mapped(%)	覆盖深度 Ave_depth	覆盖度 Cov _ratio_1X (%)	覆盖度 Cov _ratio_5X (%)	覆盖度 Cov _ratio_10X (%)
R01	103 490 495	99.6	88.91	11	97.02	80.74	55.11
R02	82 276 884	99.45	88.03	9	97.22	78.74	46.89
R03	83 316 245	99.50	87.92	9	95.40	71.64	44.06
R04	112 054 728	99.49	89.21	13	97.25	85.40	64.55

2.2 SNP的检测与注释

2.2.1 SNP检测

4个毛竹样品检测后获得SNP位点统计表（表3），其中，花毛竹样品(R04)的SNP数量最多，为1 691 715；绿皮花毛竹样品(R03)的SNP数量最少，为1 534 648。4个样品中，转换类型 (transition, Ti) SNP数量与颠换类型 (transversion, Tv) SNP数量的比值Ti/Tv为3.05-3.10之间，说明这些毛竹变型转换比颠换更容易发生。杂合类型 (heterozygosity, Het) SNP数量为纯合类型 (homozygosity, Homo) SNP数量的10倍左右，杂合比率为88.53 %~92.01 %。其中，花毛竹样品(R04)杂合比率最高，为92.01%，说明其杂合程度最高。绿皮花毛竹样品(R03)杂合比率最低，为88.53%。

表3 四个样品SNP位点统计表

Table 3 SNP loci statistics in four samples

编号 ID	SNP	转换 Transition	颠换 Transversion	转换/颠换 Ti/Tv	杂合 Het	纯合 Homo	杂合比率 Het-ratio
R01	1 628 624	1 227 674	400 950	3.06	1 476 422	152 202	90.65%
R02	1 601 748	1 207 308	394 440	3.06	1 444 235	157 513	90.16%
R03	1 534 648	1 161 232	373 416	3.10	1 358 770	175 878	88.53%
R04	1 691 715	1 274 890	416 825	3.05	1 556 713	135 002	92.01%

根据 4 个毛竹样品与参考基因组的比对结果，汇总样品间 SNP 的统计结果见表 4，表中各数值为对应的横纵两样品之间的 SNP 数。从表中可以看出，金丝毛竹(R02)与绿皮花毛竹样品(R03)间的 SNP 数最多。

表4 四个样品间的SNP统计表

Table 4 Summary of SNPs detected between four samples

编号 ID	R01	R02	R03	R04
R01	0			
R02	598 973	0		
R03	608 671	616 986	0	

R04	558 357	587 035	604 473	0

2.2.2 SNP注释

对4个样品SNP进行注释，获得其变异位点发生的区域或类型（如图1所示）。4个毛竹变型发生在CDS区域内的SNP数量占比均为2 %左右，其中，同义突变占比为48 %左右，非同义突变占比为51 %左右。非同义突变率与同义突变率的比值大于1，预示着有正向选择效应。

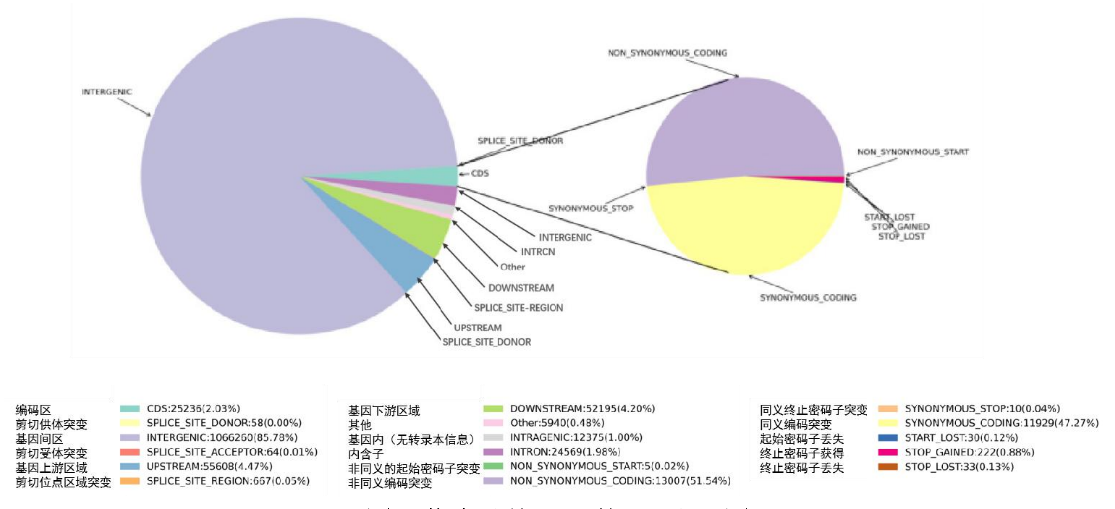


图1 黄皮毛竹(R01)的SNP注释图

Fig.1 SNP annotations pie of *Phyllostachys edulis* f. *holochrysa* (R01)

2.3 InDel检测与注释

2.3.1 InDel检测

对4个毛竹变型InDel进行统计（表5），可以发现4个样品全基因组范围检测出的InDel总数范围为271 648-292 253，其中插入类型的突变总数略低于缺失突变总数；编码区检测出的InDel总数为4 711-4 877，其中，插入突变总数为缺失突变的67 %左右。各样品中，全基因组范围内纯合突变数约为杂合突变数的2倍，编码区纯合突变数略低于杂合突变数。

表5 四个样品InDel统计表

Table 5 Summary of InDels detected in four samples

编号 ID	编码区 CDS					基因组 Genome				
	插入 Insertion	缺失 Deletion	杂合 Het	纯合 Homo	总数 Total	插入 Insertion	缺失 Deletion	杂合 Het	纯合 Homo	总数 Total
R01	2 991	1 727	2 546	2 172	4 718	139 538	143 229	94 439	188 328	282 767
R02	3 007	1 729	2 568	2 168	4 736	139 023	142 221	94 810	186 434	281 244
R03	2 996	1 715	2 572	2 139	4 711	134 641	137 007	96 117	175 531	271 648
R04	3 071	1 806	2 553	2 324	4 877	144 687	147 566	92 976	199 277	292 253

对4个样品各区域的InDel长度进行统计发现，编码区存在较多的+1、-1、+ 3、-3 类型突变，而基因组范围存在较多的+1、-1、+ 2、-2 类型突变。其中，数值代表InDel的长度(10 bp以内)；大于0为插入；小于0为缺失。

将 4 个样品的 InDel 进行两两比较，统计结果见表 6。表中各数值为对应的横纵两样品之间的 InDel 数。

表6 毛竹变型间的InDel统计表

Table 6 Summary of InDels detected between variants of Moso bamboo

编号 ID	R01	R02	R03	R04

R01	0			
R02	86 523	0		
R03	87 700	88 813	0	
R04	80 450	83 204	86 660	0

2.3.2 InDel注释

对比毛竹参考基因组的基因、CDS位置等信息，注释各样品InDel位点的发生位置、是否为移码突变等，具体注释结果如图2所示。4个毛竹变型发生在编码区的InDel数量均在1.7%左右。移码突变的InDel有可能会引起基因功能的改变。

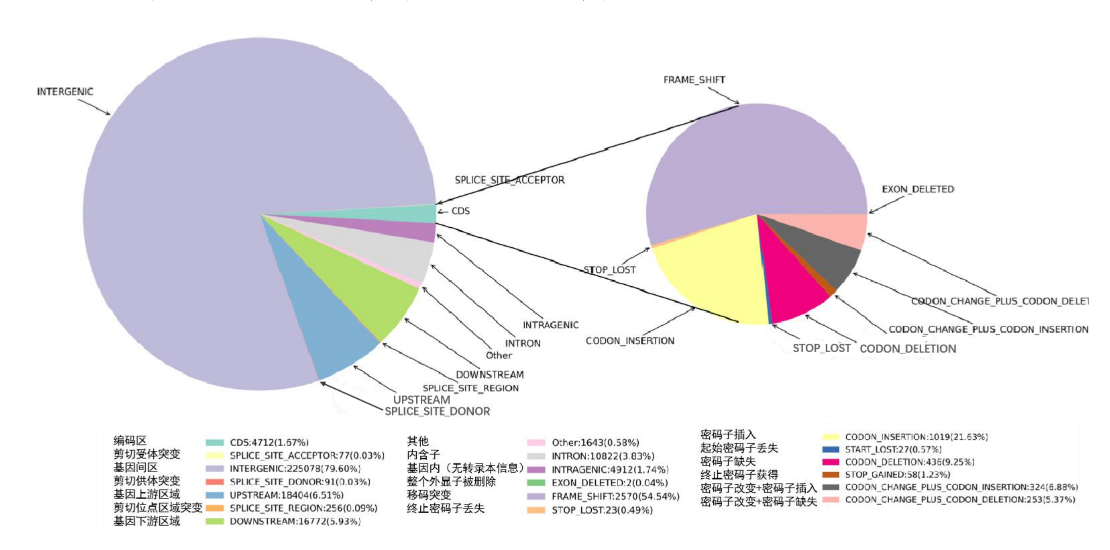


图2 黄皮毛竹(R01) InDel注释图

Fig. 2 InDel annotations pie of *Phyllostachys edulis* f. *holochrysa* (R01)

2.4 SV检测与注释

2.4.1 SV检测

检测4个样品与参考基因组间的插入(INS)、缺失(DEL)、反转(INV)、染色体内部易位(ITX)、染色体间易位(CTX)，得到的各类型SV数量统计见表7，其中，4个毛竹变型都表现为缺失类型的SV数量最多，其次为染色体内易位类型。

表7 四个样品SV数量统计

Table 7 Summary of SVs detected in four samples

编号 ID	总数 Total	插入 Insertion	缺失 Deletion	反转 Inversion	染色体内易位 Internal chromosomal translocation	染色体间易位 Translocation between chromosomes	其它 Other
R01	83 901	17 212	33 078	2 346	4 853	26 262	150
R02	77 275	15 183	32 946	1 499	4 735	22 789	123
R03	80 479	18 020	31 706	2 000	4 676	23 947	130
R04	89 754	19 541	34 613	1 166	5 133	29 142	159

2.4.2 SV注释

检测4个样品SV发生位置信息，并对缺失（DEL）、插入（INS）、反转（INV）3种类型的结构变异注释进行注释，结果（表8）表明，4个毛竹变型在各区域分布的SV总体情况一致，注释到的变异基因数目以基因间区的缺失类型最多，其次为基因间区的插入类型。

表8 四个毛竹变型结构变异注释结果统计表

Table 8 SV annotations in four variants of Moso bamboo

编号	外显子区Exon			内含子区Intron			基因间区Intergenic		
ID	缺失 Deletion	插入 Insertion	反转 Inversion	缺失 Deletion	插入 Insertion	反转 Inversion	缺失 Deletion	插入 Insertion	反转 Inversion
R01	2 258	837	237	961	389	66	29 859	15 986	2 043
R02	2 129	707	176	990	333	42	29 827	14 143	1 281
R03	2 114	953	205	942	427	44	28 650	16 640	1 751
R04	2 303	968	193	984	437	30	31 326	18 136	943

2.5 变异基因功能注释与分析

2.5.1变异基因挖掘

分别统计4个样品的非同义突变的SNP以及CDS区发生InDel和SV的基因（表9），寻找可能存在功能变异的基因。在4个毛竹样品中，花毛竹（R04）基因组存在12 555个基因变异，其中，非同义突变SNP基因为5 563个，Indel基因为4 006个，SV突变的基因为2 986个，差异基因总数和SV突变基因数最多。在3类变异基因中，非同义突变SNP基因数最多，Indel基因数量次之，SV突变的基因最少。

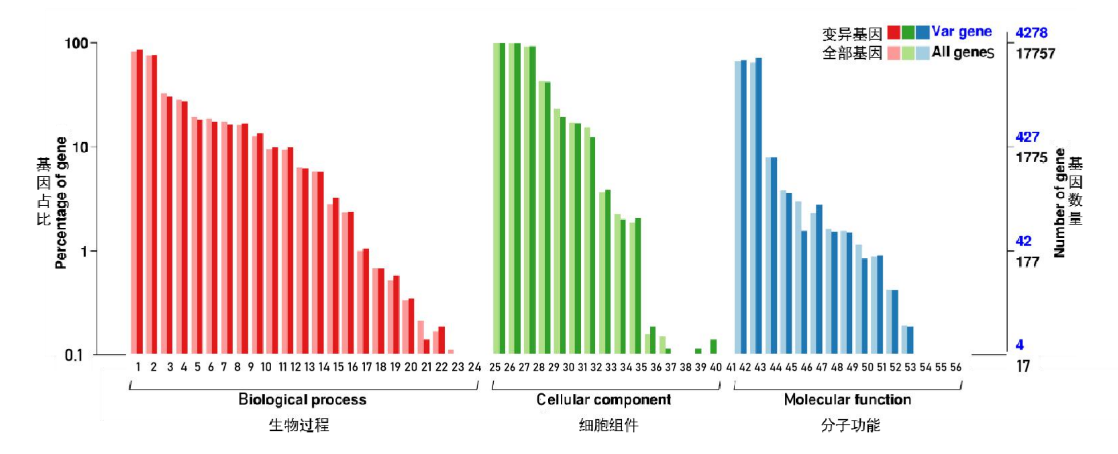
表9 四个毛竹变型的变异基因统计表

Table 9 Summary of gene variations in four variants

编号 ID	非同义突变SNP基因 Genes with Non-synonymous SNP	插入缺失基因 Genes with InDel	结构变异基因 Genes with SV	总数 Total
R01	5 405	3 905	2 903	12 213
R02	5 344	3 920	2 659	11 923
R03	5 313	3 899	2 870	12 082
R04	5 563	4 006	2 986	12 555

2.5.2变异基因的功能注释

黄皮毛竹、金丝毛竹、绿皮花毛竹和花毛竹注释到数据库中的变异基因数分别为：7 575、7 538、7 476和7 728。变异基因GO分类统计结果图（图3所示）中，显示出在3大基因功能分类体系（分子功能、细胞组件和生物过程）的56个分类内容中所对应的基因数和基因占比。其中，细胞组分分类中，与叶绿素合成相关的基因有2 431个；生物过程分类中，参与类胡萝卜素合成过程的基因有75个，与花青素合成调控以及紫外光下组织中花青素积累的相关基因有80个。4个毛竹变型在相应的基因功能中的变异基因数目有差异，例如，花毛竹有21个与类胡萝卜素合成相关的基因，而绿皮花毛竹有17个相关基因，黄皮毛竹有18个相关基因，基因数量和种类的差异，可能引起相应的功能变化。深入研究叶绿素、类胡萝卜素和花青素合成相关基因以及这些差异基因的调控途径，有利于从DNA水平上解释秆色的变异。



横坐标为 GO 的分类内容，纵坐标的左边为基因数占比，右边为基因数。 **1.** 代谢过程；**2.** 细胞过程；**3.** 对刺激的反应；**4.** 生物调节；**5.** 定位； **6.** 定位确立； **7.** 细胞组成或生物形成； **8.** 分化过程； **9.** 多细胞生物过程；**10.** 繁殖；**11.** 生殖过程；**12.** 信号；**13.** 多组织过程；**14.** 生长；**15.** 免疫系统过程；**16.** 死亡；**17.** 细胞增殖；**18.** 生物粘附；**19.** 节律过程；**20.** 病毒繁殖；**21.** 色素沉着；**22.** 运动；**23.** 细胞死亡；**24.** 碳利用；**25.** 细胞部分；**26.** 细胞；**27.** 细胞器；**28.** 膜；**29.** 细胞器部分；**30.** 膜部分；**31.** 高分子复合物；**32.** 细胞外区域；**33.** 膜内腔； **34.** 细胞连接；**35.** 细胞外基质；**36.** 类核；**37.** 病毒粒子；**38.** 细胞外基质部分；**39.** 细胞外区部分；**40.** 病毒粒子部分；**41.** 绑定；**42.** 催化活性；**43.** 运输活动；**44.** 核酸结合转录因子活性；**45.** 结构分子活性；**46.** 电子载体活性；**47.** 酶调节活性；**48.** 活动分子传感器；**49.** 抗氧化活性；**50.** 受体活性；**51.** 蛋白结合转录因子活性；**52.** 营养库活性；**53.** 翻译调节活性；**54.** 金属伴侣活性；**55.** 蛋白质标记；**56.** 通道调节活性。

Abscissa is GO classification, the left side of the ordinate is percentage of gene, and the right side is number of gene. **1.** Metabolic process; **2.** Celluar process; **3.** Response to stimulus; **4.** Biological regulation; **5.** Localization; **6.** Establishment of localization; **7.** Cellular component organization or biogenesis; **8.** Developmental process; **9.** Multicellular organismal process; **10.** Reproduction; **11.** Reproductive process; **12.** Signaling; **13.** Multi-organism process; **14.** Growth; **15.** Immune system process; **16.** Death; **17.** Cell proliferation; **18.** Biological adhesion; **19.** Rhythmic process; **20.** Viral reproduction; **21.** Pigmentation; **22.** Locomotion; **23.** Cell killing; **24.** Carbon utilization; **25.** Cell part; **26.** Cell; **27.** Organelle; **28.** Membrane; **29.** Organelle part; **30.** Membrane part; **31.** Macromolecular complex; **32.** Extracellular region; **33.** Membrane-enclosed lumen; **34.** Cell junction; **35.** Extracellular matrix; **36.** Nucleoid; **37.** Virion; **38.** Extracellular matrix part; **39.** Extracellular region part; **40.** Virion part; **41.** Binding; **42.** Catalytic activity; **43.** Transporter activity; **44.** Nucleic acid binding transcription factor activity; **45.** Structural molecule activity; **46.** Electron carrier activity; **47.** Enzyme regulator activity; **48.** Molecular transducer activity; **49.** Antioxidant activity; **50.** Receptor activity; **51.** Protein binding transcription factor activity; **52.** Nutrient reservoir activity; **53.** Translation regulator activity; **54.** Metallochaperone activity; **55.** protein tag; **56.** Channel regulator activity.

图3 黄皮毛竹(R01)变异基因的GO注释分类图

Fig. 3 Classification of *Phyllostachys edulis* f. *holochrysa* (R01) gene variations compared with GO database by blast

变异基因COG注释分类图（图4）直观显示出COG功能分类条目上分别对应的频率，其中，涉及到功能注释、转录、复制重组修复和信号转导机制的对应数值高。获得的功能注释基因1 630个，参与复制、重组和修复的基因数为369个，信号转导机制的基因数为291个，转录的相关基因222个。

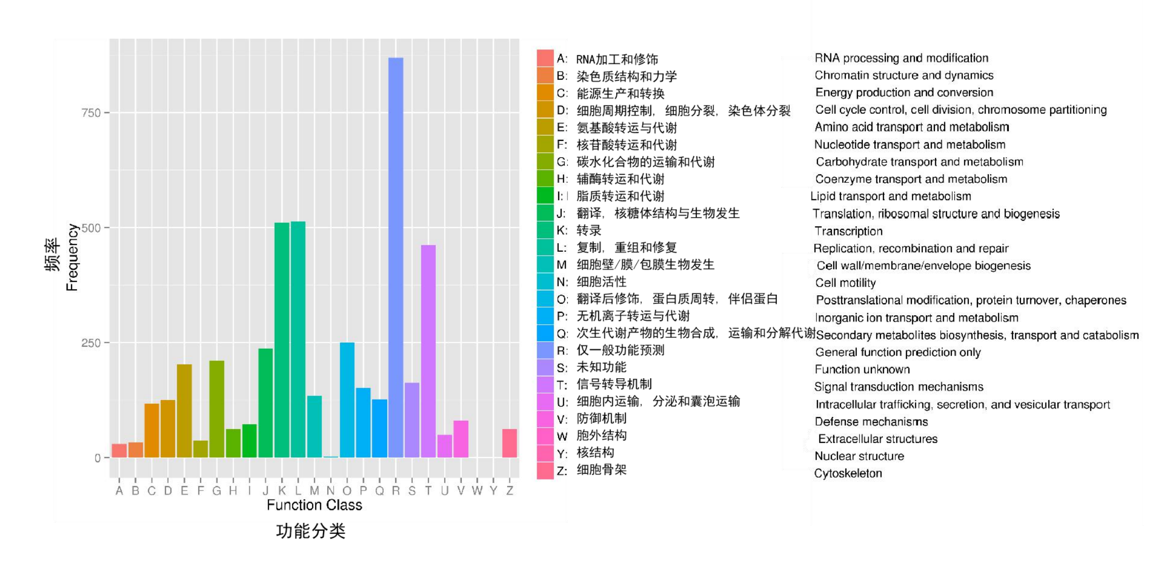


图4 黄皮毛竹(R01)变异基因的COG注释分类图

Fig. 4 Classification of *Phyllostachys edulis* f. *holochrysa* (R01) gene variations compared with COG database

KEGG数据库系统地分析4个毛竹变型的基因产物在生物学过程中的功能。以黄皮毛竹（R01）的缬氨酸、亮氨酸和异亮氨酸生物合成通路为例（图5），注释到57个基因参与该通路，其中23个变异基因。整个通路涉及不同的酶连接一系列生化反应形成，其中，框内的数字代表enzyme的号码，红色的框代表通路相关变异基因。

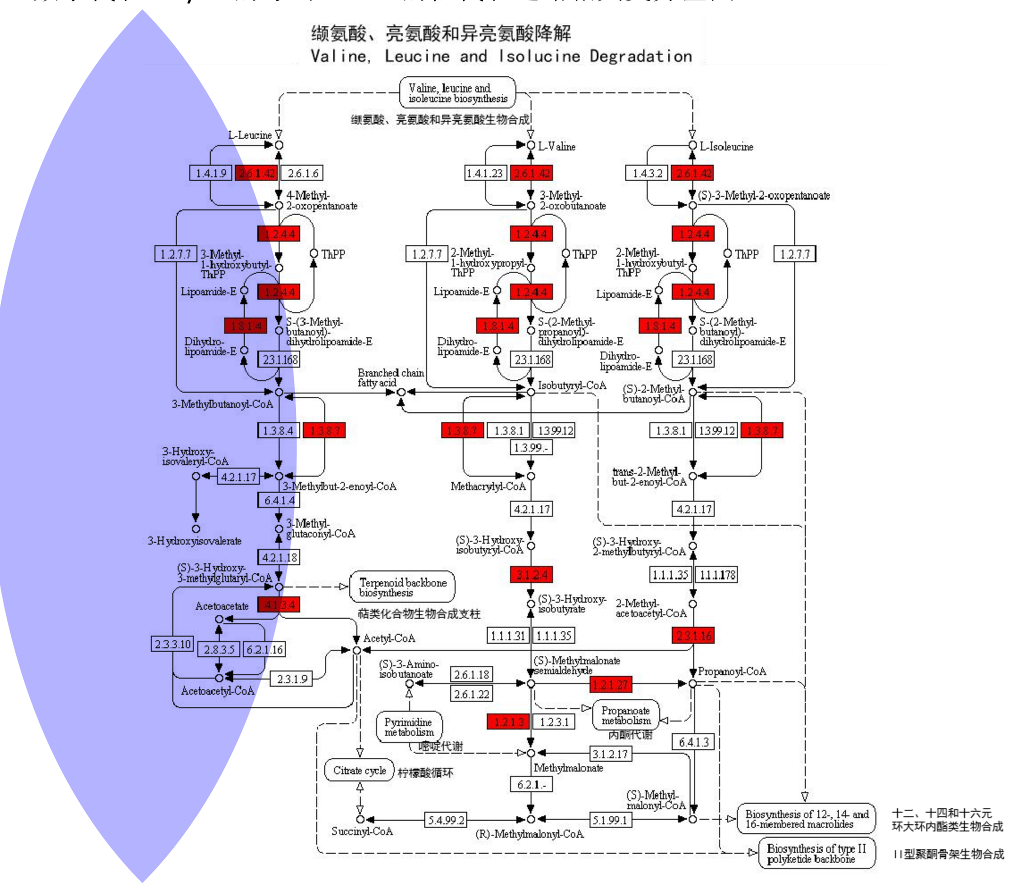


图5 黄皮毛竹(R01)变异基因的KEGG通路代谢图

Fig. 5 Pathway of *Phyllostachys edulis* f. holochrysa gene variations compared with KEGG database by blast

3 讨论与结论

3.1 讨论

全基因组重测序可以在已知植物的基因组序列基础上，对其不同品种的基因组序列进行测序，从而找出个体与该物种间的差异性（Ley et al., 2008）。随着毛竹全基因组序列的公开发表(Peng et al., 2013)，研究毛竹不同变种或变型基因组序列差异成为可能。全基因组重测序可以检测个体的全部基因组序列，扫描出一些与该个体生长性状密切相关的变异位点（宋志芳等, 2017）。毛竹的地下茎中单轴散生，其不同变异类型也都是散生竹。在毛竹自身的遗传基因漂移，以及长期的栽培措施和自然环境变迁等因素的影响下，毛竹种内产生了很多遗传变异，产生各种独特的结构形态，表现出丰富的园林观赏性状。其中，黄皮毛竹、花毛竹、绿皮花毛竹在竹秆颜色方面表现出不同程度的变异，使其具有更高的园林观赏价值。

对4个毛竹变异类型全基因组重测序，初步统计分析了其基因组数据，与毛竹参考基因组进行比对，检测其SNP、InDel和SV。SNP类型的变异分为转换和颠换两种，4个毛竹样品的转换/颠换（Ti/Tv）的比例都在3左右，说明转换类型比颠换类型更容易发生。SNP杂合比例为90%左右，说明样品有很高的杂合度，即同源染色体上SNP位点含不同类型的碱基比例高。InDel位点数同样能反映不同样品与毛竹基因组之间的差异，并且编码区的InDel会引起移码突变，影响基因功能。SV中缺失、插入、反转、易位4种类型的数量，反映出基因组水平上大片段的缺失、插入、倒置、易位等序列差异。通过生物信息学分析，比较不同秆色的变异类型在全基因组水平上的结构差异，并进行差异注释，从而为毛竹选育提供遗传基础，也为重要基因的功能研究提供有利依据。

颜色变异是植物中较常见的表型变异，其中，水稻、拟南芥、菊花等多种植物中均有叶色变异的报道。据不完全统计，水稻叶绿体含量基因超过140个（赵绍路等，2018）。竹子中的色素分为3大类：叶绿素、花青素和类胡萝卜素。通过功能数据库比对，对4个毛竹变型的变异基因，进行基因功能注释和分析。GO数据库注释聚类反映了毛竹变型在不同功能组分类中基因数目和基因产物的属性，其中，与秆色变异有关的叶绿素、类胡萝卜素和花青素等色素合成相关基因作为重点关注对象进行分析。COG数据库注释了基因产物的直系同源分类，不同分类对应的基因数目差别很大，反映了不同条件下的生理或者代谢偏好等。KEGG数据库将基因和多种酶形成通路，有氨基酸生物合成、类胡萝卜素生物合成、类黄酮生物合成、萜类化合物的生物合成、植物激素信号转导、参与卟啉和叶绿素代谢等显著富集。其中，叶绿体、类胡萝卜素和花青素等色素合成相关通路，是与秆颜色相关的主要代谢通路。

结合不同秆色毛竹变型的生物学和生理学特性，研究全基因组序列色素合成相关基因，有助于从基因水平上解析其秆色变异原因。有研究表明，毛竹不同变异类型在叶绿素含量、β胡萝卜素含量等生理指标中存在着显著性差异（陈建华等，2011），株型较大的花毛竹生理指标值比较小型的龟甲竹、绿槽毛竹大（晏育存，2011）。毛竹变型 ISSR 和 AFLP 分子标记分析表明，变型间的遗传变异程度较小（阮晓赛，2008）。在参考黄槽毛竹和黄皮花毛竹两个秆色变异毛竹变型的研究结果（牟少华等，2020）基础上，通过对黄皮毛竹等 4 个毛竹变异类型重测序，进行 DNA 水平的变异基因功能注释，可以分析基因产物在细胞中的代谢途径及功能，尤其是对黄酮类、类胡萝卜素、硝酸还原酶等合成通路的深入分析，为揭示相关代谢通路有关基因提供重要理论依据，对于探究毛竹变型秆色变异有重要意义。另外，颜色变异通常是一个不稳定的性状。例如，花毛竹在不同的生境条件下有可能变回全部绿色或者变成绿皮花毛竹，这表明竹类植物的颜色变异在遗传上不是一个稳定的性状，因此，从分子机制上探索颜色变异有其复杂性，其代谢调控还需要进一步地研究。

3.2 结论

采用第二代高通量重测序技术，对4个毛竹变型材料进行全基因组重测序研究，对其单核苷多态性（SNP）、小片段插入缺失（InDel）和结构变异（SV）进行分析和注释，筛选可能发生功能变异的基因。将变异基因与GO，COG，KEGG等功能数据库进行比对，每样品都有7 000多个变异基因得到功能注释。GO注释分类包括细胞组件、分子功能和生物过程三个基因功能分类体系的56个功能组，在细胞组分方面，叶绿素合成相关基因有2 431个；在生物过程方面，参与类胡萝卜素合成过程的基因有75个，参与花青素合成调控以及紫外光下组织中花青素积累的相关基因有80个。COG分类表明参与复制、重组和修复的基因数为369个，信号转导机制的基因数为291个，转录的相关基因222个。通过KEGG数据库系统地分析变异基因参与的黄酮类、类胡萝卜素等物质代谢合成途径。后续数据的深入分析将解析不同变异类型的基因家族和基因功能，初步阐析不同竹秆变异毛竹变型的分子遗传基础。

参考文献

ASHBURNER M, BALL C, BLAKE J. et al., 2000. Gene Ontology: tool for the unification of biology [J]. Nat Genet, (25):25-29.

BAI H, CAO YH, QUAN JZ, et al., 2013. Identifying the genome-wide sequence variations and developing new molecular markers for genetics research by resequencing a landrace cultivar of foxtail millet [J]. PLoS ONE, 8(9): 274-275.

BAI Q, HOU D, LI L, et al., 2016. Genome-wide analysis and expression characteristics of small auxin-up RNA (SAUR) genes in Moso bamboo (*Phyllostachys edulis*) [J]. Genome, 60(4): 325-336.

CHEN DM, CHEN Z, WU M, et al., 2017. Genome-wide identification and expression analysis of the HD-Zip gene family in Moso bamboo (*Phyllostachys edulis*) [J]. J Plant Growth Regul, 36(2): 323-337.

CHEN JP, YAN CY, WANG YM, et al., 2011. Physiological characteristics of 6 different variation type bamboo's leaves [J]. J of Cent S Univ of For&Tech, 31(4):70-73. [陈建华，晏存育，王艳梅，等，2011. 6 种不同变异类型毛竹叶片的生理特性研究[J]. 中南林业科技大学学报，31（4）：70-73.]

CHEN K, WALLIS JW, MCLELLAN L, et al., 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation [J]. Nat Meth, 6(9): 677-681.

CINGOLANI P, PLATTS A, WANG LL, et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3 [J]. Fly, 6(2): 80-92.

GORDON PMK, DIMNIK L, LAMONT R, et al., 2012. Optimizing genotype quality metrics for individual exomes and cohort analysis [J]. BMC Proc, 6: 21-22.

JEREMY S, CANNON SB, JESSICA S, et al., 2010. Genome sequence of the palaeopolyploid soybean [J]. Nat, 463(7278): 178-183.

JIA XP, ZHANG B, QUAN J Z, et al., 2019. Genome-wide association analysis of plant height in foxtail millet under different photoperiod conditions [J]. Acta Agric Boreai-Sin, 34(4) :16 -23. [贾小平，张博，全建章，等，2019. 不同光周期条件下谷子株高的全基因组关联分析[J]. 华北农学报，34（4）：16 -23.]

JIANG ZH, 2002. Bamboo and rattan in the world [M]. Shenyang: Liaoning Science and Technology Press: 3-5. [江泽慧，2002. 世界竹藤[M]. 沈阳：辽宁科学技术出版社：3-5.]

LEY TJ, MARDIS ER, DING L, et al., 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome[J]. Nature, 456(7218): 66-72.

LI H, DURBIN R, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. Bioinformatics, 25(14): 1754-1760.

LIN T, ZHU G, ZHANG J, et al., 2014. Genomic analyses provide insights into the history of tomato breeding [J]. Nat Genet, 46(11): 1220-1226.

LIU J, CHENG ZC, GAO J, 2019. Expression analysis and regulation network identification of the *CONSTANS-Like* gene family in Moso bamboo (*Phyllostachys edulis*) under photoperiod treatments [J]. DNA Cell Biol, 6:17-18.

MA NX, LAI GH, ZHANG PX, et al., 2014. The genus *Phyllostachys* in China [M]. Hangzhou: Zhejiang Science and Technology Press: 92-107. [马乃训，赖广辉，张培新，等，2014. 中国刚竹属[M]. 杭州：浙江科学技术出版社：92-107.]

MCKENNA A, HANNA M, BANKS E, et al., 2010. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data [J]. Genome Res, 20(9): 1297-1303.

MINORU K, SUSUMU G, SHUICHI K, et al., 2004. The KEGG resource for deciphering the genome [J]. Nucl Acid Res, 32(22): 277-280.

MU SH, LI XP, LI J, et al., 2020. Analysis on the culm shape and color variation of 4 Moso bamboo (*Phyllostachys edulis*) variants by whole genome re-sequencing [J]. Huabei Acta Agric Boreal-Sin, 35(4):1-8 [牟少华，李雪平，李娟，等，2020. 全基因组重测序分析 4 个毛竹变型的秆形和秆色变异[J]. 华北农学报，35（4）：1-8.]

PAN F, WANG Y, LIU H, et al., 2017. Genome-wide identification and expression analysis of *SBP*-like transcription factor genes in Moso bamboo (*Phyllostachys edulis*) [J]. BMC Genomics, 18(1): 486-489.

PENG ZH, LU Y, LI LB, el al, 2013. The draft genome of the fast-growing non-timber forest species Moso bamboo (*Phyllostachys heterocycla*) [J]. Nat Genet, 45(4): 456-461.

QI X, LI MW, XIE M, et al., 2014. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing [J]. Nat Comm, 5: 4340-4342.

SONG ZF, LU CL, CAO HZ, 2017. Research progress on whole genome resequencing in animal breeding [J]. Anim Husb & Vet Med, 49(11):145-148 [宋志芳，芦春莲，曹洪战，2017. 全基因组重测序及其在动物育种的研究进展[J]. 畜牧与兽医，49（11）：145-148.]

SUN H, LI L, LOU Y, et al., 2016. Genome-wide identification and characterization of aquaporin gene family in Moso bamboo (*Phyllostachys edulis*) [J]. Mol Biol Rep, 43(5): 437-441.

TAKAGI H, ABE A, YOSHIDA K, et al., 2013. QTL- seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations [J]. Plant J, 74(1):174-183.

TATUSOV RL, GALPERIN MY, NATALE DA, et al., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution [J]. Nucl Acid Res, 28(1): 33-36.

WU H, LV H, LI L, et al., 2015. Genome-wide analysis of the AP2/ERF transcription factors family and the expression patterns of DREB genes in Moso bamboo (*Phyllostachys edulis*) [J]. PLoS ONE, 10(5): e0126657.

XIE LH, LI XY, Hou D, et al., 2019. Genome-wide analysis and expression profiling of the heat shock factor gene family in *Phyllostachys edulis* during development and in response to abiotic stresses [J]. Forests, 10(2):100-105.

YAN CY, 2011. Different variation type's bamboo with bamboo'S biology physiology characteristic [D]. Changsha: Central South University of Forestry&Technology:39-43. [晏育存，2011. 毛竹不同变异类型的生物学与生理学特性研究[D]. 长沙：中南林业科技大学：39-43.]

YIN YP, DING QJ, LUO JW, et al., 2020. Genomic sequencing analysis of *Magonolia officinalis* based on Pacbio's third-generation sequencing technology.[尹彦棚，丁乔娇，罗加伟等，2020. 基于 Pacbio 第三代测序技术的厚朴基因组测序分析[J]. http://kns.cnki.net/kcms/detail/45.1134.Q.20200415.1147.008.html.]

YUAN XS, 2008. Genetic variations of *Phyllostachys edulis* provenance and cultivars by ISSR and AFLP [D]. Hangzhou: Zhejiang Forestry College:30-31. [阮晓赛，2008. 毛竹种源及栽培变种遗传变异的 AFLP 和 ISSR 分析[D]. 杭州：浙江林学院：30-31.]

ZHANG SP, QIU SL, ZHENG YY, et al., 2017. The purple *Abelmoschus esculentus* transcriptome as a source for

gene sequence information [J]. J of Nucl Agric Sci 15(4): 643-653. [张少平，邱珊莲，郑云云，等，2017. 紫色黄秋葵转录组功能基因测序及分析[J]. 核农学报，15（4）：643-653.]

ZHANG YW，LI W, ZHANG LF, et al., 2016. Genome-wide variations of soybean cultivar Qihuang 34 by whole genome resequencing [J]. Chin J Oil Crop Sci, 38(2):150-158. [张彦威，李伟，张礼凤，等，2016. 基于重测序的大豆新品种齐黄 34 的全基因组变异挖掘[J]. 中国油料作物学报，38（2）：150-158.]

ZHAO SL, LIU K, WAN BJ, et al., 2018. Advances in research on rice leaf color mutants[J/OL]. Barley and Cereal Sci, 35(6):1-6. [赵绍路，刘凯，宛柏杰，等，2018. 水稻叶色突变研究进展[J/OL]. 大麦与谷类科学，35（6）：l-6.]

ZHOU Z, JIANG Y, WANG Z, et al., 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean [J]. Nat Biotechnol, 33(4): 408-414.

ZIDANI S, FERCHICHI A, CHAIEB M, 2005. Genomic DNA extraction method from pearl millet (*Pennisetum glaucum*) leaves [J]. Afr J Biotechnol, 4: 862-866.